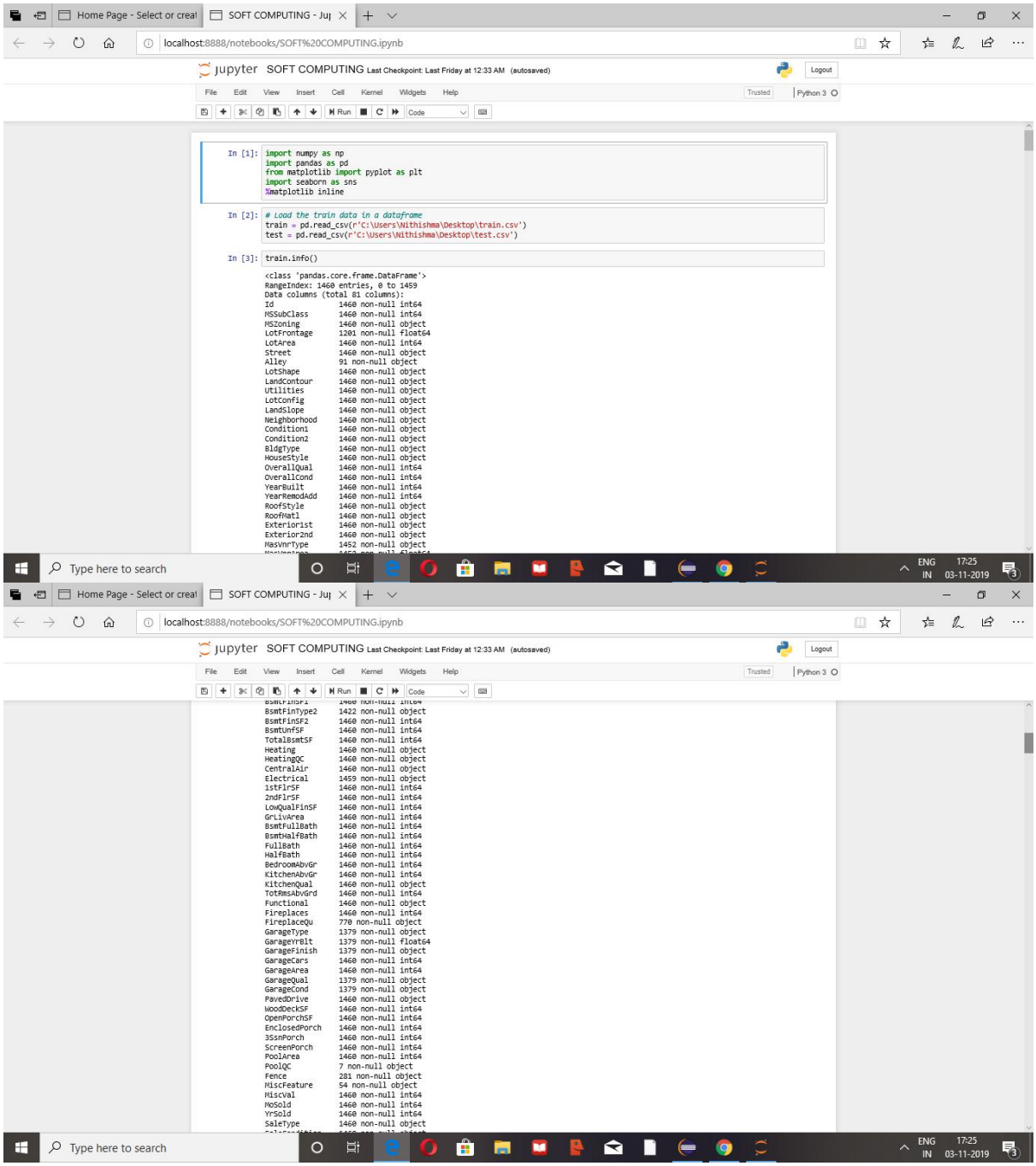
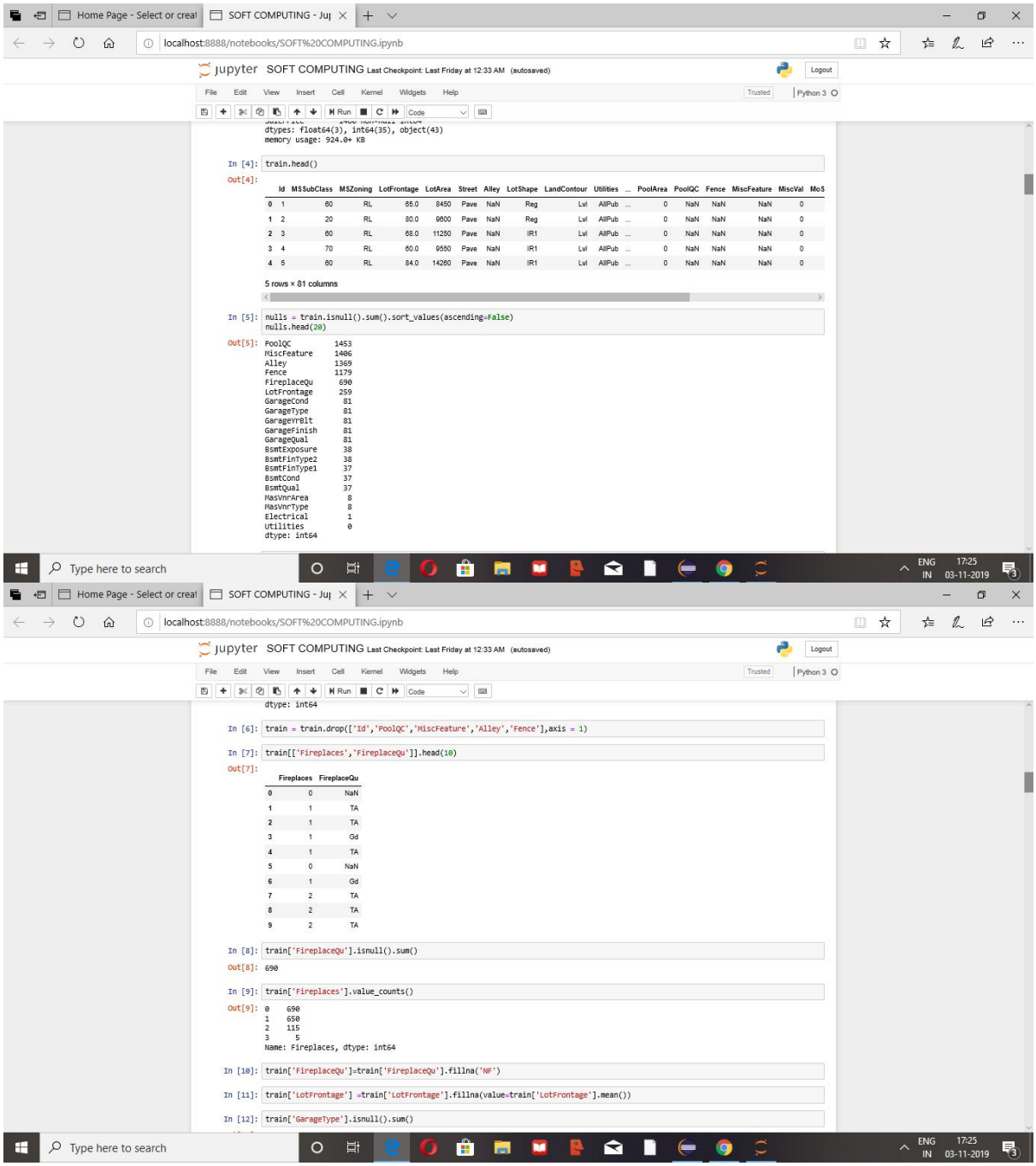
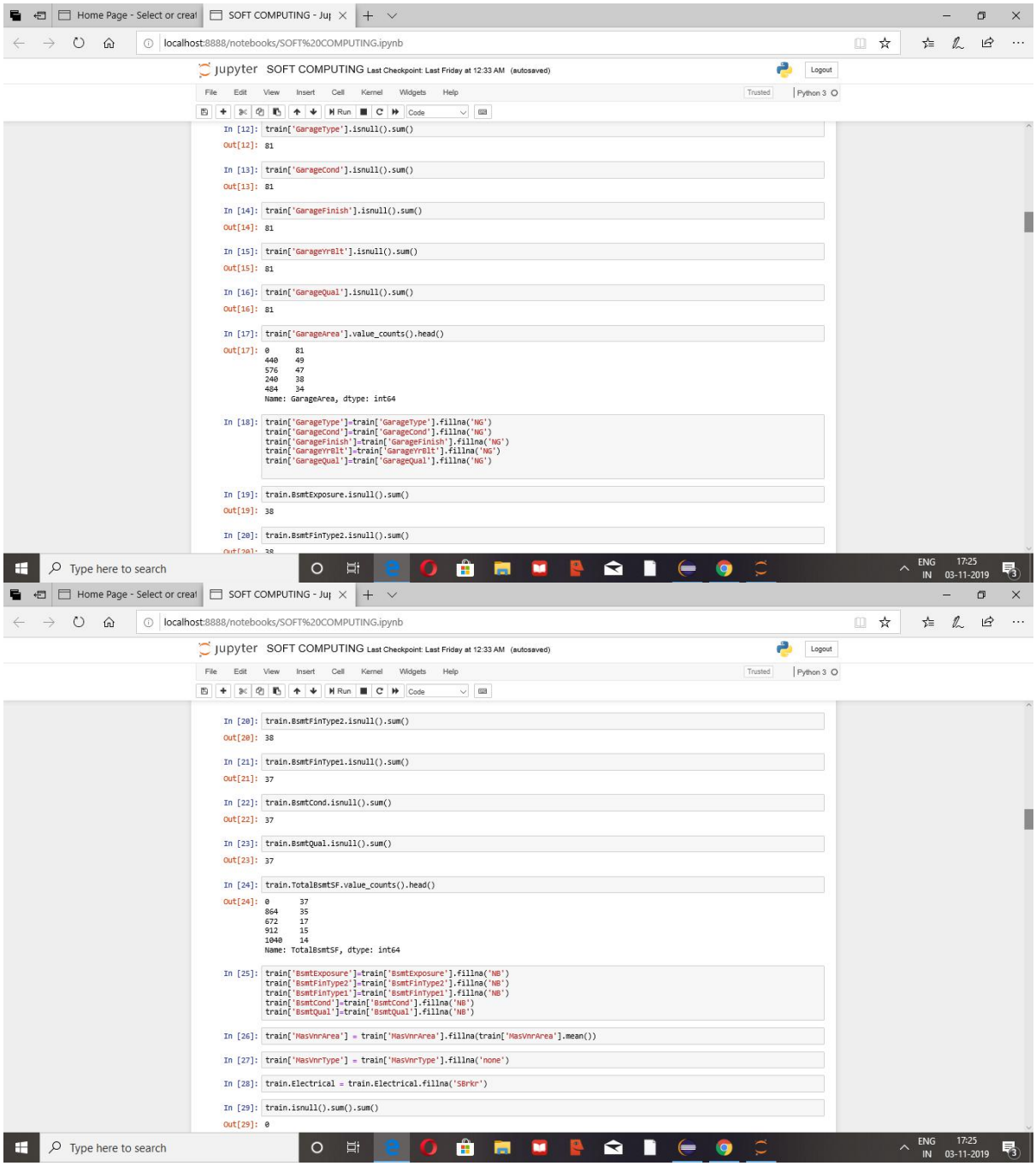

Home Page - Select or creat
SOFT COMPUTING - Jup
localhost:8888/notebooks/SOFT%20COMPUTING.ipynb
jupyter SOFT COMPUTING Last Checkpoint: Last Friday at 12:33 AM (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
In [1]:
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
import seaborn as sns
%matplotlib inline
In [2]:
# Load the train data in a dataframe
train = pd.read_csv(r'C:\Users\Nithishma\Desktop\train.csv')
test = pd.read_csv(r'C:\Users\Nithishma\Desktop\test.csv')
In [3]:
train.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1460 entries, 0 to 1459
Data columns (total 81 columns):
Id 1460 non-null int64
MSSubClass 1460 non-null int64
MSZoning 1460 non-null object
LotFrontage 1201 non-null float64
LotArea 1460 non-null int64
Street 1460 non-null object
Alley 91 non-null object
LotShape 1460 non-null object
LandContour 1460 non-null object
Utilities 1460 non-null object
LotConfig 1460 non-null object
LandSlope 1460 non-null object
Neighborhood 1460 non-null object
Condition1 1460 non-null object
Condition2 1460 non-null object
BldgType 1460 non-null object
HouseStyle 1460 non-null object
OverallQual 1460 non-null int64
OverallCond 1460 non-null int64
YearBuilt 1460 non-null int64
YearRemodAdd 1460 non-null int64
RoofStyle 1460 non-null object
RoofMatl 1460 non-null object
Exterior1st 1460 non-null object
Exterior2nd 1460 non-null object
MasVnrType 1452 non-null object
Type here to search
ENG
IN
17:25
03-11-2019

Home Page - Select or creat
SOFT COMPUTING - Jup
localhost:8888/notebooks/SOFT%20COMPUTING.ipynb
jupyter SOFT COMPUTING Last Checkpoint: Last Friday at 12:33 AM (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
BsmtFinSF1 1460 non-null int64
BsmtFinType2 1422 non-null object
BsmtFinSF2 1460 non-null int64
BsmtUnfSF 1460 non-null int64
TotalBsmtSF 1460 non-null int64
Heating 1460 non-null object
HeatingQC 1460 non-null object
CentralAir 1460 non-null object
Electrical 1459 non-null object
1stFlrSF 1460 non-null int64
2ndFlrSF 1460 non-null int64
LowQualFinSF 1460 non-null int64
GrLivArea 1460 non-null int64
BsmtFullBath 1460 non-null int64
BsmtHalfBath 1460 non-null int64
FullBath 1460 non-null int64
HalfBath 1460 non-null int64
BedroomAbvGr 1460 non-null int64
KitchenAbvGr 1460 non-null int64
KitchenQual 1460 non-null object
TotRmsAbvGrd 1460 non-null int64
Functional 1460 non-null object
Fireplaces 1460 non-null int64
FireplaceQu 770 non-null object
GarageType 1379 non-null object
GarageYrBlt 1379 non-null float64
GarageFinish 1379 non-null object
GarageCars 1460 non-null int64
GarageArea 1460 non-null int64
GarageQual 1379 non-null object
GarageCond 1379 non-null object
PavedDrive 1460 non-null object
WoodDeckSF 1460 non-null int64
OpenPorchSF 1460 non-null int64
EnclosedPorch 1460 non-null int64
3SsnPorch 1460 non-null int64
ScreenPorch 1460 non-null int64
PoolArea 1460 non-null int64
PoolQC 7 non-null object
Fence 281 non-null object
MiscFeature 54 non-null object
MiscVal 1460 non-null int64
MoSold 1460 non-null int64
YrSold 1460 non-null int64
SaleType 1460 non-null object
Type here to search
ENG
IN
17:25
03-11-2019


Home Page - Select or creat
SOFT COMPUTING - Jup
localhost:8888/notebooks/SOFT%20COMPUTING.ipynb
jupyter SOFT COMPUTING Last Checkpoint: Last Friday at 12:33 AM (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
dtypes: float64(3), int64(35), object(43)
memory usage: 924.0+ KB
In [4]:
train.head()
Out[4]:
Id MSSubClass MSZoning LotFrontage LotArea Street Alley LotShape LandContour Utilities ... PoolArea PoolQC Fence MiscFeature MiscVal MoS
0 1 60 RL 65.0 8450 Pave NaN Reg Lvl AllPub ... 0 NaN NaN NaN 0
1 2 20 RL 80.0 9600 Pave NaN Reg Lvl AllPub ... 0 NaN NaN NaN 0
2 3 60 RL 68.0 11250 Pave NaN IR1 Lvl AllPub ... 0 NaN NaN NaN 0
3 4 70 RL 60.0 9550 Pave NaN IR1 Lvl AllPub ... 0 NaN NaN NaN 0
4 5 60 RL 84.0 14260 Pave NaN IR1 Lvl AllPub ... 0 NaN NaN NaN 0
5 rows × 81 columns
In [5]:
nulls = train.isnull().sum().sort_values(ascending=False)
nulls.head(20)
Out[5]:
PoolQC 1453
MiscFeature 1406
Alley 1369
Fence 1179
FireplaceQu 690
LotFrontage 259
GarageCond 81
GarageType 81
GarageYrBlt 81
GarageFinish 81
GarageQual 81
BsmtExposure 38
BsmtFinType2 38
BsmtFinType1 37
BsmtCond 37
BsmtQual 37
MasVnrArea 8
MasVnrType 8
Electrical 1
Utilities 0
dtype: int64
Type here to search
ENG
IN
17:25
03-11-2019
Home Page - Select or creat
SOFT COMPUTING - Jup
localhost:8888/notebooks/SOFT%20COMPUTING.ipynb
jupyter SOFT COMPUTING Last Checkpoint: Last Friday at 12:33 AM (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
dtype: int64
In [6]:
train = train.drop(['Id','PoolQC','MiscFeature','Alley','Fence'],axis = 1)
In [7]:
train[['Fireplaces','FireplaceQu']].head(10)
Out[7]:
Fireplaces FireplaceQu
0 0 NaN
1 1 TA
2 1 TA
3 1 Gd
4 1 TA
5 0 NaN
6 1 Gd
7 2 TA
8 2 TA
9 2 TA
In [8]:
train['FireplaceQu'].isnull().sum()
Out[8]:
690
In [9]:
train['Fireplaces'].value_counts()
Out[9]:
0 690
1 650
2 115
3 5
Name: Fireplaces, dtype: int64
In [10]:
train['FireplaceQu']=train['FireplaceQu'].fillna('NF')
In [11]:
train['LotFrontage'] =train['LotFrontage'].fillna(value=train['LotFrontage'].mean())
In [12]:
train['GarageType'].isnull().sum()
Type here to search
ENG
IN
17:25
03-11-2019


Home Page - Select or creat
SOFT COMPUTING - Jup
localhost:8888/notebooks/SOFT%20COMPUTING.ipynb
jupyter SOFT COMPUTING Last Checkpoint: Last Friday at 12:33 AM (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
In [12]: train['GarageType'].isnull().sum()
Out[12]: 81
In [13]: train['GarageCond'].isnull().sum()
Out[13]: 81
In [14]: train['GarageFinish'].isnull().sum()
Out[14]: 81
In [15]: train['GarageYrBlt'].isnull().sum()
Out[15]: 81
In [16]: train['GarageQual'].isnull().sum()
Out[16]: 81
In [17]: train['GarageArea'].value_counts().head()
Out[17]: 0 81
440 49
576 47
240 38
484 34
Name: GarageArea, dtype: int64
In [18]: train['GarageType']=train['GarageType'].fillna('NG')
train['GarageCond']=train['GarageCond'].fillna('NG')
train['GarageFinish']=train['GarageFinish'].fillna('NG')
train['GarageYrBlt']=train['GarageYrBlt'].fillna('NG')
train['GarageQual']=train['GarageQual'].fillna('NG')
In [19]: train.BsmtExposure.isnull().sum()
Out[19]: 38
In [20]: train.BsmtFinType2.isnull().sum()
Type here to search
ENG
IN
17:25
03-11-2019

Home Page - Select or creat
SOFT COMPUTING - Jup
localhost:8888/notebooks/SOFT%20COMPUTING.ipynb
jupyter SOFT COMPUTING Last Checkpoint: Last Friday at 12:33 AM (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
In [20]: train.BsmtFinType2.isnull().sum()
Out[20]: 38
In [21]: train.BsmtFinType1.isnull().sum()
Out[21]: 37
In [22]: train.BsmtCond.isnull().sum()
Out[22]: 37
In [23]: train.BsmtQual.isnull().sum()
Out[23]: 37
In [24]: train.TotalBsmtSF.value_counts().head()
Out[24]: 0 37
864 35
672 17
912 15
1040 14
Name: TotalBsmtSF, dtype: int64
In [25]: train['BsmtExposure']=train['BsmtExposure'].fillna('NB')
train['BsmtFinType2']=train['BsmtFinType2'].fillna('NB')
train['BsmtFinType1']=train['BsmtFinType1'].fillna('NB')
train['BsmtCond']=train['BsmtCond'].fillna('NB')
train['BsmtQual']=train['BsmtQual'].fillna('NB')
In [26]: train['MasVnrArea'] = train['MasVnrArea'].fillna(train['MasVnrArea'].mean())
In [27]: train['MasVnrType'] = train['MasVnrType'].fillna('none')
In [28]: train.Electrical = train.Electrical.fillna('SBrkr')
In [29]: train.isnull().sum().sum()
Out[29]: 0
Type here to search
ENG
IN
17:25
03-11-2019

Home Page - Select or creat | SOFT COMPUTING - Jup

localhost:8888/notebooks/SOFT%20COMPUTING.ipynb

jupyter SOFT COMPUTING Last Checkpoint: Last Friday at 12:33 AM (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3

Run Code

In [30]:
```
num_train = train._get_numeric_data()
```

In [31]:
```
num_train.columns
```

Out[31]:
```
Index(['MSSubClass', 'LotFrontage', 'LotArea', 'OverallQual', 'OverallCond',
       'YearBuilt', 'YearRemodAdd', 'MasVnrArea', 'BsmtFinSF1', 'BsmtFinSF2',
       'BsmtUnfSF', 'TotalBsmtSF', '1stFlrSF', '2ndFlrSF', 'LowQualFinSF',
       'GrLivArea', 'BsmtFullBath', 'BsmtHalfBath', 'FullBath', 'HalfBath',
       'BedroomAbvGr', 'KitchenAbvGr', 'TotRmsAbvGrd', 'Fireplaces',
       'GarageCars', 'GarageArea', 'WoodDeckSF', 'OpenPorchSF',
       'EnclosedPorch', '3SsnPorch', 'ScreenPorch', 'PoolArea', 'MiscVal',
       'MoSold', 'YrSold', 'SalePrice'],
      dtype='object')
```

In [32]:
```
def var_summary(x):
    return pd.Series([x.count(), x.isnull().sum(), x.sum(), x.mean(), x.median(),  x.std(), x.var(), x.min(), x.quantile(0.01), 
                      index=['N', 'NMISS', 'SUM', 'MEAN','MEDIAN', 'STD', 'VAR', 'MIN', 'P1' , 'P5' ,'P10' ,'P25' ,'P50' ,'P75' ,'P9

num_train.apply(lambda x: var_summary(x)).T
```

Out[32]:

	N	NMISS	SUM	MEAN	MEDIAN	STD	VAR	MIN	P1	P5	P10	P25
MSSubClass	1460.0	0.0	8.307000e+04	56.897260	50.000000	42.300571	1.789338e+03	20.0	20.00	20.00	20.0	20.00
LotFrontage	1460.0	0.0	1.022729e+05	70.049958	70.049958	22.024023	4.850576e+02	21.0	21.00	35.95	49.0	60.00
LotArea	1460.0	0.0	1.535457e+07	10516.828082	9478.500000	9981.264932	9.962585e+07	1300.0	1680.00	3311.70	5000.0	7553.50
OverallQual	1460.0	0.0	8.905000e+03	6.099315	6.000000	1.382997	1.912679e+00	1.0	3.00	4.00	5.0	5.00
OverallCond	1460.0	0.0	8.140000e+03	5.575342	5.000000	1.112799	1.238322e+00	1.0	3.00	4.00	5.0	5.00
YearBuilt	1460.0	0.0	2.878051e+06	1971.267808	1973.000000	30.202904	9.122154e+02	1872.0	1899.18	1916.00	1924.9	1954.00
YearRemodAdd	1460.0	0.0	2.897904e+06	1984.865753	1994.000000	20.645407	4.262328e+02	1950.0	1950.00	1950.00	1950.0	1967.00
MasVnrArea	1460.0	0.0	1.513805e+05	103.685262	0.000000	180.569112	3.260520e+04	0.0	0.00	0.00	0.0	0.00
BsmtFinSF1	1460.0	0.0	6.477140e+05	443.639726	383.500000	456.098091	2.080255e+05	0.0	0.00	0.00	0.0	0.00
BsmtFinSF2	1460.0	0.0	6.796200e+04	46.549315	0.000000	161.319273	2.602391e+04	0.0	0.00	0.00	0.0	0.00
BsmtUnfSF	1460.0	0.0	8.281710e+05	567.240411	477.500000	441.866955	1.952464e+05	0.0	0.00	0.00	74.9	223.00
TotalBsmtSF	1460.0	0.0	1.543847e+06	1057.429452	991.500000	438.705324	1.924624e+05	0.0	0.00	519.30	636.9	795.75
1stFlrSF	1460.0	0.0	1.697435e+06	1162.626712	1087.000000	386.587738	1.494501e+05	334.0	520.00	672.95	756.9	882.00
2ndFlrSF	1460.0	0.0	5.068090e+05	346.992466	0.000000	436.528436	1.905571e+05	0.0	0.00	0.00	0.0	0.00

Type here to search

ENG IN 17:25 03-11-2019

Home Page - Select or creat | SOFT COMPUTING - Jup

localhost:8888/notebooks/SOFT%20COMPUTING.ipynb

jupyter SOFT COMPUTING Last Checkpoint: Last Friday at 12:33 AM (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

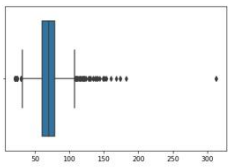
Trusted | Python 3

Run Code

BedroomAbvGr	1460.0	0.0	4.185000e+03	2.866438	3.000000	0.815778	6.654998e-01	0.0	1.00	2.00	2.0	2.00
KitchenAbvGr	1460.0	0.0	1.528000e+03	1.046575	1.000000	0.220338	4.854892e-02	0.0	1.00	1.00	1.0	1.00
TotRmsAbvGrd	1460.0	0.0	9.516000e+03	6.517808	6.000000	1.625393	2.641903e+00	2.0	3.00	4.00	5.0	5.00
Fireplaces	1460.0	0.0	8.950000e+02	0.613014	1.000000	0.644666	4.155947e-01	0.0	0.00	0.00	0.0	0.00
GarageCars	1460.0	0.0	2.580000e+03	1.767123	2.000000	0.747315	5.584797e-01	0.0	0.00	0.00	1.0	1.00
GarageArea	1460.0	0.0	6.905510e+05	472.980137	480.000000	213.804841	4.571251e+04	0.0	0.00	0.00	240.0	334.50
WoodDeckSF	1460.0	0.0	1.375970e+05	94.244521	0.000000	125.338794	1.570981e+04	0.0	0.00	0.00	0.0	0.00
OpenPorchSF	1460.0	0.0	6.812400e+04	46.660274	25.000000	66.256028	4.389861e+03	0.0	0.00	0.00	0.0	0.00
EnclosedPorch	1460.0	0.0	3.205300e+04	21.954110	0.000000	61.119149	3.735550e+03	0.0	0.00	0.00	0.0	0.00
3SsnPorch	1460.0	0.0	4.978000e+03	3.409589	0.000000	29.317331	8.595059e+02	0.0	0.00	0.00	0.0	0.00
ScreenPorch	1460.0	0.0	2.198900e+04	15.060959	0.000000	55.757415	3.108889e+03	0.0	0.00	0.00	0.0	0.00
PoolArea	1460.0	0.0	4.028000e+03	2.758904	0.000000	40.177307	1.614216e+03	0.0	0.00	0.00	0.0	0.00
MiscVal	1460.0	0.0	6.349400e+04	43.489041	0.000000	496.123024	2.461381e+05	0.0	0.00	0.00	0.0	0.00
MoSold	1460.0	0.0	9.230000e+03	6.321918	6.000000	2.703626	7.309595e+00	1.0	1.00	2.00	3.0	5.00
YrSold	1460.0	0.0	2.931411e+06	2007.815753	2008.000000	1.328095	1.763837e+00	2006.0	2006.00	2006.00	2006.0	2007.00
SalePrice	1460.0	0.0	2.641449e+08	180921.195890	163000.000000	79442.502883	6.311111e+09	34900.0	61815.97	88000.00	106475.0	129975.00

In [33]:
```
sns.boxplot([num_train.LotFrontage])
```

Out[33]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1caad0f08d0>
```



Type here to search

ENG IN 17:25 03-11-2019

Home Page - Select or creat | SOFT COMPUTING - Jup

localhost:8888/notebooks/SOFT%20COMPUTING.ipynb

jupyter SOFT COMPUTING Last Checkpoint: Last Friday at 12:33 AM (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

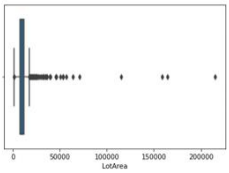
Trusted | Python 3

Run Code

In [62]:
```
train['LotFrontage']= train['LotFrontage'].clip(upper=train['LotFrontage'].quantile(0.99))
```

In [35]:
```
sns.boxplot(num_train.LotArea)
```

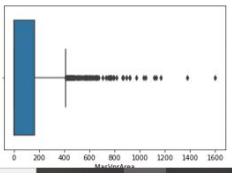
Out[35]: <matplotlib.axes._subplots.AxesSubplot at 0x1caa6351748>



In [63]:
```
train['LotArea']= train['LotArea'].clip(upper=train['LotArea'].quantile(0.99))
```

In [37]:
```
sns.boxplot(train['MasVnrArea'])
```

Out[37]: <matplotlib.axes._subplots.AxesSubplot at 0x1caad518208>



Type here to search

ENG IN 17:26 03-11-2019

Home Page - Select or creat | SOFT COMPUTING - Jup

localhost:8888/notebooks/SOFT%20COMPUTING.ipynb

jupyter SOFT COMPUTING Last Checkpoint: Last Friday at 12:33 AM (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

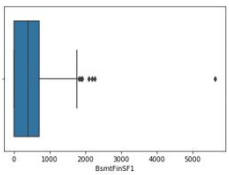
Trusted | Python 3

Run Code

In [64]:
```
train['MasVnrArea']= train['MasVnrArea'].clip(upper=train['MasVnrArea'].quantile(0.99))
```

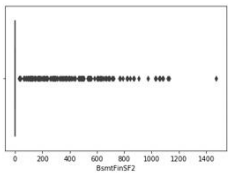
In [39]:
```
sns.boxplot(train['BsmtFinSF1'])
```

Out[39]: <matplotlib.axes._subplots.AxesSubplot at 0x1caad551908>



In [40]:
```
sns.boxplot(train['BsmtFinSF2'])
```

Out[40]: <matplotlib.axes._subplots.AxesSubplot at 0x1caad5ec4a8>



In [61]:
```
train['BsmtFinSF1']= train['BsmtFinSF1'].clip(upper=train['BsmtFinSF1'].quantile(0.99))
```

Type here to search

ENG IN 17:26 03-11-2019

Home Page - Select or creat | SOFT COMPUTING - Jup

localhost:8888/notebooks/SOFT%20COMPUTING.ipynb

jupyter SOFT COMPUTING Last Checkpoint: Last Friday at 12:33 AM (autosaved)

Logout

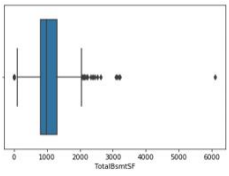
File Edit View Insert Cell Kernel Widgets Help | Trusted | Python 3

Run | Code

In [61]:
```
train['BsmtFinSF1']= train['BsmtFinSF1'].clip(upper=train['BsmtFinSF1'].quantile(0.99))
train['BsmtFinSF2']= train['BsmtFinSF2'].clip(upper=train['BsmtFinSF2'].quantile(0.99))
```

In [42]:
```
sns.boxplot(train['TotalBsmtSF'])
```

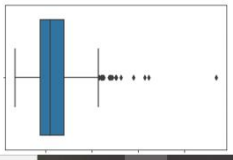
Out[42]: <matplotlib.axes._subplots.AxesSubplot at 0x1caad5eca58>



In [60]:
```
train['TotalBsmtSF']= train['TotalBsmtSF'].clip(upper=train['TotalBsmtSF'].quantile(0.99))
```

In [44]:
```
sns.boxplot(train['1stFlrSF'])
```

Out[44]: <matplotlib.axes._subplots.AxesSubplot at 0x1caad6b8d30>



Type here to search

ENG IN 17:26 03-11-2019

Home Page - Select or creat | SOFT COMPUTING - Jup

localhost:8888/notebooks/SOFT%20COMPUTING.ipynb

jupyter SOFT COMPUTING Last Checkpoint: Last Friday at 12:33 AM (autosaved)

Logout

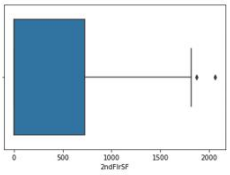
File Edit View Insert Cell Kernel Widgets Help | Trusted | Python 3

Run | Code

In [65]:
```
train['1stFlrSF']= train['1stFlrSF'].clip(upper=train['1stFlrSF'].quantile(0.99))
```

In [46]:
```
sns.boxplot(train['2ndFlrSF'])
```

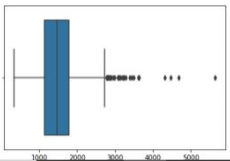
Out[46]: <matplotlib.axes._subplots.AxesSubplot at 0x1caad715080>



In [66]:
```
train['2ndFlrSF']= train['2ndFlrSF'].clip(upper=train['2ndFlrSF'].quantile(0.99))
```

In [48]:
```
sns.boxplot(train['GrLivArea'])
```

Out[48]: <matplotlib.axes._subplots.AxesSubplot at 0x1caad77a320>



Type here to search

ENG IN 17:26 03-11-2019

Home Page - Select or creat | SOFT COMPUTING - Jup

localhost:8888/notebooks/SOFT%20COMPUTING.ipynb

jupyter SOFT COMPUTING Last Checkpoint: Last Friday at 12:33 AM (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3

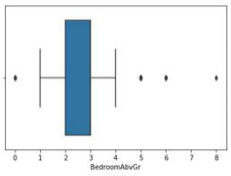
Run Code

1000 2000 3000 4000 5000
GrLivArea

In [67]:
```
train['GrLivArea']= train['GrLivArea'].clip(upper=train['GrLivArea'].quantile(0.99))
```

In [50]:
```
sns.boxplot(train['BedroomAbvGr'])
```

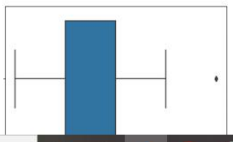
Out[50]: <matplotlib.axes._subplots.AxesSubplot at 0x1caad7d67b8>



In [68]:
```
train['BedroomAbvGr']= train['BedroomAbvGr'].clip(upper=train['BedroomAbvGr'].quantile(0.99))
train['BedroomAbvGr']= train['BedroomAbvGr'].clip(lower=train['BedroomAbvGr'].quantile(0.01))
```

In [52]:
```
sns.boxplot(train['GarageCars'])
```

Out[52]: <matplotlib.axes._subplots.AxesSubplot at 0x1caad832470>



Type here to search

ENG IN 17:26 03-11-2019

Home Page - Select or creat | SOFT COMPUTING - Jup

localhost:8888/notebooks/SOFT%20COMPUTING.ipynb

jupyter SOFT COMPUTING Last Checkpoint: Last Friday at 12:33 AM (autosaved)

Logout

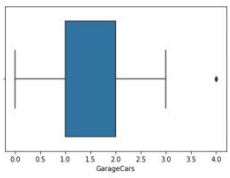
File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3

Run Code

In [52]:
```
sns.boxplot(train['GarageCars'])
```

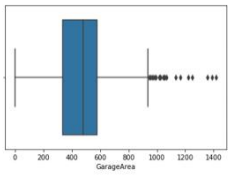
Out[52]: <matplotlib.axes._subplots.AxesSubplot at 0x1caad832470>



In [69]:
```
train['GarageCars']= train['GarageCars'].clip(upper=train['GarageCars'].quantile(0.99))
```

In [54]:
```
sns.boxplot(train['GarageArea'])
```

Out[54]: <matplotlib.axes._subplots.AxesSubplot at 0x1caad8281d0>



In [70]:
```
train['GarageArea']= train['GarageArea'].clip(upper=train['GarageArea'].quantile(0.99))
```

Type here to search

ENG IN 17:26 03-11-2019

Home Page - Select or creat | SOFT COMPUTING - Jup

localhost:8888/notebooks/SOFT%20COMPUTING.ipynb

jupyter SOFT COMPUTING Last Checkpoint: Last Friday at 12:33 AM (autosaved)

Logout

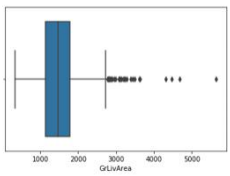
File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3

Run Code

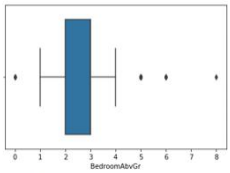
```
In [66]: train['2ndFlrSF']= train['2ndFlrSF'].clip(upper=train['2ndFlrSF'].quantile(0.99))

In [48]: sns.boxplot(train['GrLivArea'])
Out[48]: <matplotlib.axes._subplots.AxesSubplot at 0x1caad77a320>
```




```
In [67]: train['GrLivArea']= train['GrLivArea'].clip(upper=train['GrLivArea'].quantile(0.99))

In [50]: sns.boxplot(train['BedroomAbvGr'])
Out[50]: <matplotlib.axes._subplots.AxesSubplot at 0x1caad7d67b8>
```



Type here to search

ENG IN 17:28 03-11-2019

Home Page - Select or creat | SOFT COMPUTING - Jup

localhost:8888/notebooks/SOFT%20COMPUTING.ipynb

jupyter SOFT COMPUTING Last Checkpoint: Last Friday at 12:33 AM (autosaved)

Logout

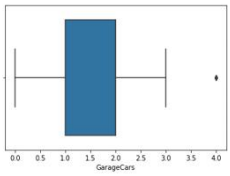
File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3

Run Code

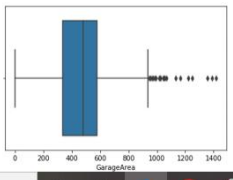
```
In [68]: train['BedroomAbvGr']= train['BedroomAbvGr'].clip(upper=train['BedroomAbvGr'].quantile(0.99))
         train['BedroomAbvGr']= train['BedroomAbvGr'].clip(lower=train['BedroomAbvGr'].quantile(0.01))

In [52]: sns.boxplot(train['GarageCars'])
Out[52]: <matplotlib.axes._subplots.AxesSubplot at 0x1caad832470>
```




```
In [69]: train['GarageCars']= train['GarageCars'].clip(upper=train['GarageCars'].quantile(0.99))

In [54]: sns.boxplot(train['GarageArea'])
Out[54]: <matplotlib.axes._subplots.AxesSubplot at 0x1caad8281d0>
```



Type here to search

ENG IN 17:28 03-11-2019

Home Page - Select or creat | SOFT COMPUTING - Jup

localhost:8888/notebooks/SOFT%20COMPUTING.ipynb

jupyter SOFT COMPUTING Last Checkpoint: Last Friday at 12:33 AM (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

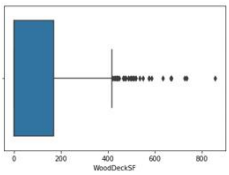
Trusted | Python 3

Run Code

In [70]:
```
train['GarageArea']= train['GarageArea'].clip(upper=train['GarageArea'].quantile(0.99))
```

In [56]:
```
sns.boxplot(train['WoodDeckSF'])
```

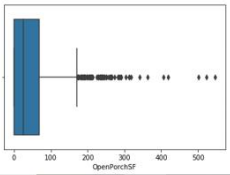
Out[56]: <matplotlib.axes._subplots.AxesSubplot at 0x1caad907e48>



In [71]:
```
train['WoodDeckSF']= train['WoodDeckSF'].clip(upper=train['WoodDeckSF'].quantile(0.99))
```

In [58]:
```
sns.boxplot(train['OpenPorchSF'])
```

Out[58]: <matplotlib.axes._subplots.AxesSubplot at 0x1caad96f7b8>



Type here to search

ENG IN 17:28 03-11-2019

Home Page - Select or creat | SOFT COMPUTING - Jup

localhost:8888/notebooks/SOFT%20COMPUTING.ipynb

jupyter SOFT COMPUTING Last Checkpoint: Last Friday at 12:33 AM (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

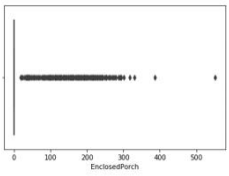
Trusted | Python 3

Run Code

In [72]:
```
train['OpenPorchSF']= train['OpenPorchSF'].clip(upper=train['OpenPorchSF'].quantile(0.99))
```

In [73]:
```
sns.boxplot(train['EnclosedPorch'])
```

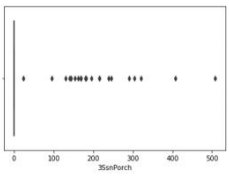
Out[73]: <matplotlib.axes._subplots.AxesSubplot at 0x1caad9b8f98>



In [74]:
```
train['EnclosedPorch']= train['EnclosedPorch'].clip(upper=train['EnclosedPorch'].quantile(0.99))
```

In [75]:
```
sns.boxplot(train['3SsnPorch'])
```

Out[75]: <matplotlib.axes._subplots.AxesSubplot at 0x1caada1fda0>



Type here to search

ENG IN 17:28 03-11-2019

Home Page - Select or creat | SOFT COMPUTING - Jup

localhost:8888/notebooks/SOFT%20COMPUTING.ipynb

jupyter SOFT COMPUTING Last Checkpoint: Last Friday at 12:33 AM (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

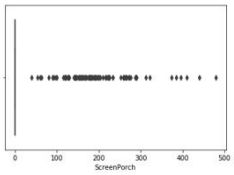
Trusted | Python 3

Run Code

In [76]: train['3SsnPorch']= train['3SsnPorch'].clip(upper=train['3SsnPorch'].quantile(0.99))

In [77]:
```
sns.boxplot(train['ScreenPorch'])
```

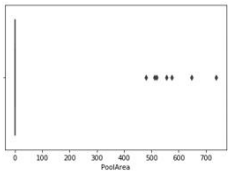
Out[77]: <matplotlib.axes._subplots.AxesSubplot at 0x1caada87a20>



In [78]:
```
train['ScreenPorch']= train['ScreenPorch'].clip(upper=train['ScreenPorch'].quantile(0.99))
```

In [79]:
```
sns.boxplot(train['PoolArea'])
```

Out[79]: <matplotlib.axes._subplots.AxesSubplot at 0x1caadaf3828>



Type here to search

ENG IN 17:28 03-11-2019

Home Page - Select or creat | SOFT COMPUTING - Jup

localhost:8888/notebooks/SOFT%20COMPUTING.ipynb

jupyter SOFT COMPUTING Last Checkpoint: Last Friday at 12:33 AM (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

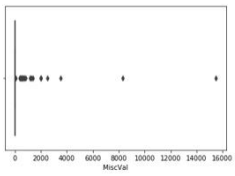
Trusted | Python 3

Run Code

In [80]:
```
train['PoolArea']= train['PoolArea'].clip(upper=train['PoolArea'].quantile(0.99))
```

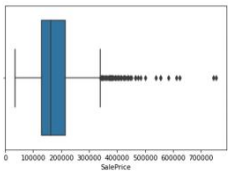
In [81]:
```
sns.boxplot(train['MiscVal'])
```

Out[81]: <matplotlib.axes._subplots.AxesSubplot at 0x1caadb52128>



In [82]:
```
sns.boxplot(train.SalePrice)
```

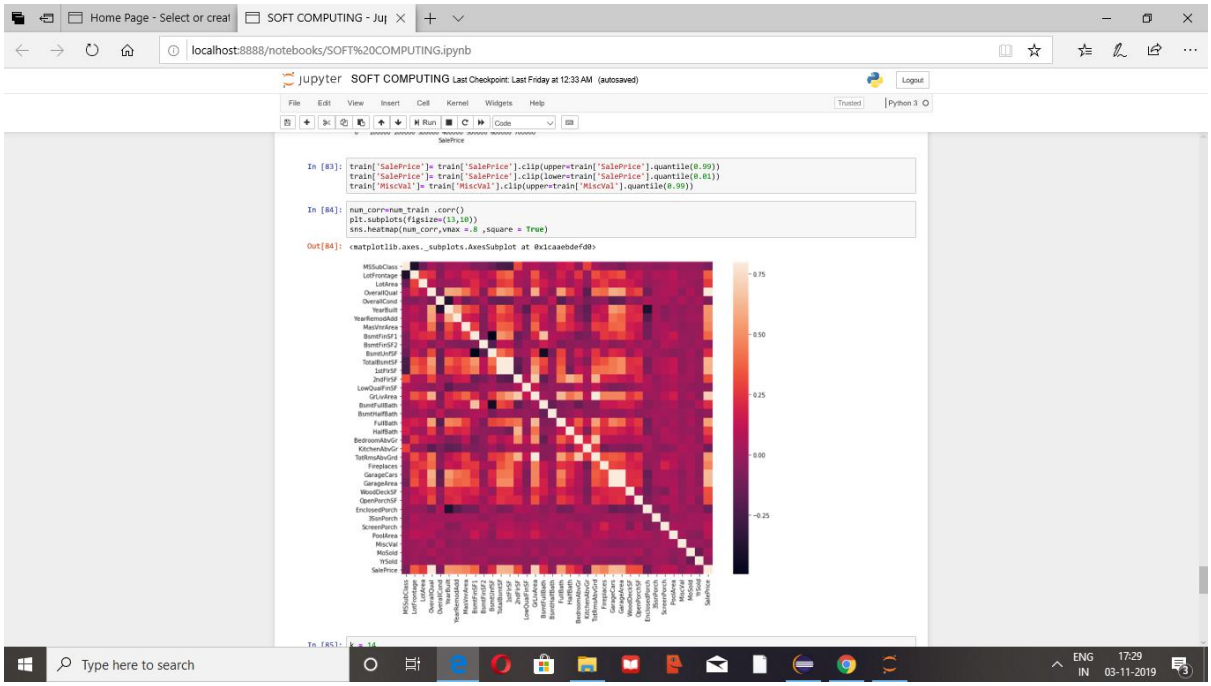
Out[82]: <matplotlib.axes._subplots.AxesSubplot at 0x1caadbbb2e8>



In [83]: train['SalePrice']= train['SalePrice'].clip(upper=train['SalePrice'].quantile(0.99))

Type here to search

ENG IN 17:28 03-11-2019


Home Page - Select or creat
SOFT COMPUTING - Jup
localhost:8888/notebooks/SOFT%20COMPUTING.ipynb
jupyter SOFT COMPUTING Last Checkpoint: Last Friday at 12:33 AM (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
SalePrice
In [83]:
train['SalePrice']= train['SalePrice'].clip(upper=train['SalePrice'].quantile(0.99))
train['SalePrice']= train['SalePrice'].clip(lower=train['SalePrice'].quantile(0.01))
train['MiscVal']= train['MiscVal'].clip(upper=train['MiscVal'].quantile(0.99))
In [84]:
num_corr=num_train .corr()
plt.subplots(figsize=(13,10))
sns.heatmap(num_corr,vmax =.8 ,square = True)
Out[84]:
<matplotlib.axes._subplots.AxesSubplot at 0x1caaebdefd0>
MSSubClass
LotFrontage
LotArea
OverallQual
OverallCond
YearBuilt
YearRemodAdd
MasVnrArea
BsmtFinSF1
BsmtFinSF2
BsmtUnfSF
TotalBsmtSF
1stFlrSF
2ndFlrSF
LowQualFinSF
GrLivArea
BsmtFullBath
BsmtHalfBath
FullBath
HalfBath
BedroomAbvGr
KitchenAbvGr
TotRmsAbvGrd
Fireplaces
GarageCars
GarageArea
WoodDeckSF
OpenPorchSF
EnclosedPorch
3SsnPorch
ScreenPorch
PoolArea
MiscVal
MoSold
YrSold
SalePrice
0.75
0.50
0.25
0.00
−0.25
In [85]:
k = 14
Type here to search
ENG
IN
17:29
03-11-2019

Home Page - Select or creat | SOFT COMPUTING - Jup

localhost:8888/notebooks/SOFT%20COMPUTING.ipynb

jupyter SOFT COMPUTING Last Checkpoint: Last Friday at 12:33 AM (autosaved)

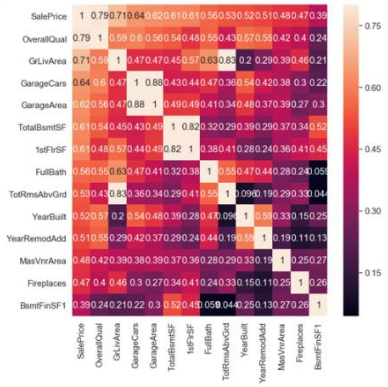
Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3

Run Code

In [85]:
```
k = 14
cols = num_corr.nlargest(k, 'SalePrice')['SalePrice'].index
cm = np.corrcoef(num_train[cols].values.T)
sns.set(font_scale=1.35)
f, ax = plt.subplots(figsize=(10,10))
hm=sns.heatmap(cm, annot = True,vmax =.8, yticklabels=cols.values, xticklabels = cols.values)
```



In [86]:
```
from sklearn.preprocessing import StandardScaler
```

Type here to search

ENG IN 17:29 03-11-2019

Home Page - Select or creat | SOFT COMPUTING - Jup

localhost:8888/notebooks/SOFT%20COMPUTING.ipynb

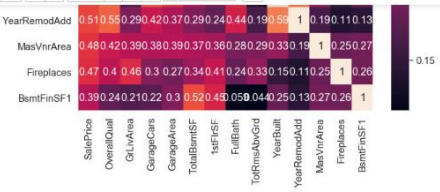
jupyter SOFT COMPUTING Last Checkpoint: Last Friday at 12:33 AM (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3

Run Code



In [86]:
```
from sklearn.preprocessing import StandardScaler
```

In [87]:
```
train_d = pd.get_dummies(train)
train_d1 = train_d.drop(['SalePrice'],axis = 1)
y = train_d.SalePrice
```

In [88]:
```
scaler = StandardScaler()
scaler.fit(train_d1)
t_train = scaler.transform(train_d1)
```

In [89]:
```
from sklearn.decomposition import PCA
pca_hp = PCA(30)
x_fit = pca_hp.fit_transform(t_train)
```

In [91]:
```
np.exp(pca_hp.explained_variance_ratio_)
```

Out[91]:
```
array([1.05208629, 1.02448825, 1.01999132, 1.0184108 , 1.0169433 ,
       1.01510818, 1.01251127, 1.0122223 , 1.0099264 , 1.0097077 ,
       1.00938108, 1.0086314 , 1.00824231, 1.00812493, 1.00790935,
       1.00767959, 1.00745176, 1.00731303, 1.00721918, 1.00688895,
       1.00679322, 1.00677786, 1.00655226, 1.00643388, 1.00625254,
       1.00624528, 1.00615529, 1.00603923, 1.00588235, 1.00583934])
```

In [ ]:

Type here to search

ENG IN 17:29 03-11-2019